American Journal of Industrial Medicine 13:119–130 (1988)

Indirect Corrections for Confounding Under Multiplicative and Additive Risk Models

M.H. Gail, MD, PhD, S. Wacholder, PhD, and J.H. Lubin, PhD

We define a multiplicative model and an additive model for the hazards associated jointly with exposure and with the presence of a confounder like smoking. Under the multiplicative model, the crude relative risk may be adjusted indirectly, by means of a factor proposed by Axelson [1978], and implicitly by Cornfield et al. [1959] and Schlesselman [1978]. We present corresponding indirect correction formulas under the additive risk model for the risk difference and for the excess relative risk. Conditions are established under which these corrections may be applied to age-adjusted rates from composite study populations. We demonstrate that indirect corrections may be no better than crude measures of risk if one assumes the wrong model for the joint action of the exposure and confounding factors. These results are illustrated on an example of occupational exposure to vermiculite. The limitations of the techniques are discussed.

Key words: relative risk, standardized mortality ratio, risk difference, excess relative risk, cohort studies

INTRODUCTION

A recent conference entitled "Obtaining and using information on smoking in occupational epidemiologic studies" included a workshop on "indirect methods" for adjusting relative risk estimates in the absence of specific confounder (smoking) data for individuals. The basic strategy is to adjust crude relative risk measures by using external information on the joint distribution of confounder and exposure, together with external information on the relative risk of disease due to the confounding factor among unexposed individuals. The paper by Axelson and Steenland [1988] highlights and studies these methods.

We now explore the relationship between underlying models for the joint risk of exposure and confounding factors and the "indirect" methods used to adjust crude risk estimates. Lubin and Gaffey [1988] discuss multiplicative relative risk models as well as additive relative risk models for the joint action of exposure and confounding factors. We first define the multiplicative model and consider how to adjust the crude relative risk, which is appropriate under a multiplicative risk model. Next we define the additive risk model and show how to adjust the crude risk difference and crude

Epidemiologic Methods Section, Biostatistics Branch, National Cancer Institute, Bethesda, Maryland.
Address reprint requests to M.H. Gail, MD, PhD, Epidemiologic Methods Section, Biostatistics Branch, National Cancer Institute, Landow Building, Room 3C37, Bethesda MD 20892.
Accepted for publication July 28, 1987.

excess relative risk, which are appropriate under an additive risk model. We then apply these ideas to occupational studies in which data on individual smoking habits are not available and consider the data of Amandus and Wheeler [1987] on lung cancer risk among vermiculite miners.

In the last three sections we discuss the consequences of mistakenly applying the wrong adjustment, as might happen if the wrong model is assumed. Just as using the wrong model with complete information on exposure and confounding can lead to misleading summary measures of risk from exposure, so can indirect adjustments that are based on the wrong model be misleading. Moreover, indirect adjustments are performed in the absence of data on confounding for individuals; in this setting we have no internal information with which to select a model for joint action. A previously reported study may provide guidance in the selection of a model for the joint risks of exposure and confounder, and such information can be crucial in making valid indirect adjustments.

THE MULTIPLICATIVE MODEL AND INDIRECT CORRECTION FOR THE HAZARD RATIO

Axelson [1978] first proposed an indirect method to correct for confounding in relative hazards, namely, the ratios of incidence rates. This method yields valid estimates of relative risk if the effects of the confounder and the exposure act multiplicatively on the risk, as we discuss below. Similar factors are given by Cornfield et al. [1959], Miettinen [1972], and Schlesselman [1978] for relative risks, namely, ratios of probabilities of events, without regard to time. We shall follow the approach of Axelson [1978] by expressing our results in terms of hazard rates (events per person-year).

Let I_{ijt} be the hazard (incidence or mortality) rate per person-year at age t for an individual at exposure level i (i = 1 if exposed, i = 0 otherwise) and confounder level j (j = 1 if the confounder is present and j = 0 otherwise). For the moment we suppress the notation t, but we are always considering a specific age group, t. Later, we make simplifying assumptions to allow one to carry out indirect corrections even when age-specific data on the risk of exposure and confounding factors are not available. The multiplicative model is

$$\begin{aligned} I_{01} &= RR_C I_{00}, \\ I_{10} &= RR_E I_{00}, \\ \text{and } I_{11} &= I_{10} I_{01}/I_{00} \\ &= (I_{10}/I_{00})(I_{01}/I_{00})I_{00} \\ &\equiv (RR_E)(RR_C)I_{00}. \end{aligned} \tag{1}$$

This model implies that the relative hazard of exposure is constant over levels of the confounder and that the relative hazard of the confounder is constant over levels of exposure. Under the multiplicative model we seek a valid estimate of RR_E by correcting the crude incidence ratio,

$$\begin{aligned}RR_E^{CRUDE} &\equiv \{\pi_1 I_{11} + (1 - \pi_1)I_{10}\}/\{\pi_0 I_{01} + (1 - \pi_0)I_{00}\} \\ &= (I_{10}/I_{00})\, \{\pi_1 RR_C + (1 - \pi_1)\}/\{\pi_0 RR_C + (1 - \pi_0)\} \\ &= RR_E BIAS_M,\end{aligned} \tag{2}$$

where $\pi_1 = P\,(j = 1\,|\,i = 1)$ and $\pi_0 = P\,(j = 1\,|\,i = 0)$ are the proportions of subjects at age t with the confounding factor respectively among those exposed and those not exposed. There is no confounding ($BIAS_M = 1$) either if $\pi_1 = \pi_0$ or if $RR_C = 1$. If the quantities RR_C, π_1, and π_0 are known, then a proper correction to RR_E^{CRUDE} may be made by dividing by $BIAS_M$. These results agree with Axelson [1978].

To apply this correction to composite data from different age groups, one would need to have the correct values of $RR_C(t)$, $\pi_1(t)$, and $\pi_0(t)$ for each age, and one would need to assume a proportional hazards model of the form

$$I_{11t}/I_{01t} = I_{10t}/I_{00t} = RR_E, \tag{3}$$

where RR_E does not depend on age. The relative risk from confounding, $RR_C(t) = I_{01t}/I_{00t}$ and the proportions with the confounding factor, $\pi_1(t)$ and $\pi_0(t)$ may still depend on age under this model. For example, the proportion of an unexposed population that smokes, $\pi_0(t)$, depends on age. Therefore a rigorous correction procedure would obtain corrected estimates of RR_E separately for each age group and then take a weighted average of these corrected relative hazards. We call model 3 a "stratified proportional hazards model" because it only requires constancy of RR_E over age within strata defined by levels of the confounder.

It might be argued that age-specific information on $RR_C(t)$, $\pi_0(t)$, and $\pi_1(t)$ are usually unavailable. Then the simple procedures used by Axelson [1978] may be justified under the assumptions that RR_C, π_1, and π_0 do not depend on age. The assumptions that RR_E and RR_C are independent of t imply that I_{11t}/I_{00t}, I_{10t}/I_{00t}, and I_{01t}/I_{00t} are free of t, even though I_{00t} may vary with age. We call this a "full proportional hazards model," because hazard ratios for all levels of the exposure and confounder variables must be constant over age. Under these assumptions, the correction factor, $BIAS_M$, does not depend on age, and it may be demonstrated (see Appendix) that:

1. The ratio of the age-adjusted rates for the exposed and unexposed cohorts is (RR_E) $(BIAS_M)$, and therefore the same correction factor, $BIAS_M$, for the confounder may be used for the composite age-adjusted rates as for a single age stratum. All this depends on the strong assumptions above, but this result holds whether direct or indirect age adjustment is used.
2. The age-adjusted SMR for the exposed cohort, computed with respect to a general unexposed population with known age-specific rates, satisfies

$$SMR = (RR_E)\,(BIAS_M),$$

where RR_E is the ratio of rates in the exposed and general reference populations. This ratio is assumed to be the same for all subgroups defined by age and the confounder. Again, the same correction factor $BIAS_M$ may be used, provided π_0 is known for the general population.

The argument is analogous for case-control data. Indeed, case-control data stratified for age will yield an age-adjusted common relative hazard [Prentice and Breslow, 1978] or RR_E^{CRUDE}, which may be divided by $BIAS_M$ to estimate RR_E.

We note that the use of t to index age may be broadened to index a cross-classification on age, calendar year, and other factors that might typically be controlled by direct or indirect standardization. The assumption that RR_E is constant over t, as in equation 3, or the assumption that a full proportional hazards model holds true, would then extend to each cell indexed by t in such a cross-classification. Similar generalizations of the use of t apply to equation 6 and to the "additive hazards model" discussed directly below.

THE ADDITIVE MODEL AND INDIRECT CORRECTION FOR THE EXCESS RISK AND EXCESS RELATIVE RISK

The additive risk model at age t is given by

$$\begin{aligned} I_{01} &= I_{00} + RD_C, \\ I_{10} &= I_{00} + RD_E, \\ \text{and } I_{11} &= RD_C + RD_E + I_{00} \\ &= I_{01} + I_{10} - I_{00}, \end{aligned} \tag{4}$$

where t is suppressed. We seek a valid estimate of RD_E, the risk difference. The crude risk difference

$$\begin{aligned} RD_E^{CRUDE} &= \{\pi_1(I_{10} + I_{01} - I_{00}) + (1 - \pi_1)I_{10}\} \\ &\quad - \{\pi_0 I_{01} + (1 - \pi_0)I_{00}\} \\ &= (I_{10} - I_{00}) + (\pi_1 - \pi_0)(I_{01} - I_{00}) \\ &= RD_E + (\pi_1 - \pi_0)RD_C \\ &\equiv RD_E + BIAS_A. \end{aligned} \tag{5}$$

As before, the crude risk difference is a valid estimate of RD_E (i.e., $BIAS_A = 0$) if $RD_C = 0$ or if $\pi_1 = \pi_0$. Otherwise an appropriate correction is made by subtracting $BIAS_A(t)$ from $RD_E^{CRUDE}(t)$ for those in age group t.

To treat composite populations composed of multiple age groups, we suppose

$$I_{11t} - I_{01t} = I_{10t} - I_{00t} = RD_E \tag{6}$$

is constant over age groups t, which is analogous to equation 3. As before, we suppose in general that $\pi_1(t)$, $\pi_0(t)$, and $RD_C(t) = I_{01t} - I_{00t}$ depend on age, so that $BIAS_A(t)$ also depends on age. Thus, in general, a separate correction should be made for each age group, and a weighted average of these corrected estimates RD_E taken as an estimate of RD_E.

Again, if age-specific information is not available, indirect adjustments are feasible under the simplifying assumptions that π_1, π_0, and RD_C are independent of age. Then $BIAS_A$ is also independent of age. The assumptions that RD_E and RD_C are independent of t imply that $I_{11t} - I_{00t}$, $I_{10t} - I_{00t}$, and $I_{01t} - I_{00t}$ are independent of t, even though I_{00t} may depend on t. These relations are analogous to the proportional hazards model and might be termed the "additive hazards model." Under these assumptions, it is shown in the Appendix that any *direct adjustment procedure* that assigns the same age distribution to both exposure groups yields a crude risk difference of $RD_E + BIAS_A$, so that the common bias term $BIAS_A$ may be subtracted to correct the direct age-adjusted estimate of risk difference.

Although the risk difference RD_E is an important measure of the public health burden imposed by exposure, this quantity cannot be estimated from case-control studies, which only yield data on relative hazards [Miettinen, 1976; Prentice and Breslow, 1978]. Under the additive model 4, at a fixed age t, one can still examine the excess relative risk from exposure,

$$ERR_E = (I_{10} - I_{00})/I_{00} = (I_{11} - I_{01})/I_{00}.$$

Note that ERR_E is expressible as the difference of two relative risks and can therefore be estimated from case-control data. Dividing the additive model 4 by I_{00}, we reparameterize it in terms of ERR_E and

$$\begin{aligned} ERR_C &= (I_{01} - I_{00})/I_{00} = (I_{11} - I_{10})/I_{00} \\ \text{as } I_{01} &= I_{00}(ERR_C + 1), \\ I_{10} &= I_{00}(ERR_E + 1), \\ \text{and } I_{11} &= I_{00}(ERR_E + ERR_C + 1). \end{aligned} \tag{7}$$

These relationships are equivalent to the definitions of additivity and excess relative risk in equations 15-2, 15-3, and 15-4 of Rothman [1986]. For a fixed age group, t, we can thereby re-express the crude excess relative risk as

$$\begin{aligned} ERR_E^{CRUDE} &= RD_E^{CRUDE}/\{I_{01}\pi_0 + I_{00}(1 - \pi_0)\} \\ &= (RD_E + BIAS_A)/\{I_{01}\pi_0 + I_{00}(1 - \pi_0)\} \\ &= I_{00}\{ERR_E + (\pi_1 - \pi_0)ERR_C\}/\{I_{01}\pi_0 + I_{00}(1 - \pi_0)\} \\ &= \{ERR_E + (\pi_1 - \pi_0)ERR_C\}/\{ERR_C\pi_0 + 1\}. \end{aligned} \tag{8}$$

Note that $ERR_E^{CRUDE} = ERR_E$ if $ERR_C = 0$, namely if C is not an independent risk factor. However, unlike previous cases, $\pi_1 = \pi_0$ is not sufficient to ensure $ERR_E^{CRUDE} = ERR_E$. Only when $ERR_E = 0$ does $\pi_1 = \pi_0$ ensure no confounding.

To correct ERR_C, we solve 8 to obtain

$$ERR_E = ERR_E^{CRUDE}\{ERR_C\pi_0 + 1\} - (\pi_1 - \pi_0)ERR_C. \quad (9$$

This correction applies to a specific age group t.

Suppose ERR_E is constant over age groups. Then separate corrected estimate of ERR_E may be obtained for each age group, provided $\pi_0(t)$, $\pi_1(t)$, and $ERR_C(t)$ ar known, and an overall estimate of ERR_E is obtained as a weighted average o corrected age-specific estimates.

As before, if age-specific data are not available, we can proceed under th assumptions that $\pi_1(t)$, $\pi_0(t)$, and $ERR_C(t)$ are independent of time. The assumption that ERR_E and ERR_C are both independent of time imply the "proportional hazard model," namely I_{11t}/I_{00t}, I_{10t}/I_{00t}, and I_{01t}/I_{00t} are independent of time, although I_{0c} may vary with time. The absolute rates I_{11t}, I_{10t}, and I_{01t} are given by equation 7 Under these assumptions, it is demonstrated in the Appendix that the same correctio equation 9 may be applied to the composite age-adjusted excess relative risk as th excess relative risk for any single age stratum. However, the age adjustment shoul be based on direct, and not indirect standardization for age.

Under the same assumptions, a case-control study with stratification on ag will yield an age-adjusted estimate of the relative hazard for exposure that equal RR_E^{CRUDE}, from which $ERR_E^{CRUDE} = RR_E^{CRUDE} - 1$ may be computed. The quantit ERR_E^{CRUDE} may be corrected via equation 9 to obtain a valid estimate of ERR_E.

The definition of excess relative risk discussed in this section is commonly use and is constant over levels of the confounder under an additive risk model. Under th multiplicative risk model, one could define $ERR_E = (I_{10} - I_{00})/I_{00} = (I_{11} - I_{01})/I_{0}$ instead. Then, from the relationship $ERR_E = RR_E - 1$, one would obtain the indirec correction $ERR_E = (ERR_E^{CRUDE} + 1)(BIAS_M) - 1$. In the remainder of this paper however, we adhere to the definition and methods appropriate for the additive ris model.

EXAMPLE

Amandus and Wheeler [1987] observed 20 cases of lung cancer during 13,50 person-years of follow-up in a cohort of vermiculite miners. They calculated 8.9 lung cancer deaths would be "expected" from general U.S. age-specific lung cance death rates. If one sets the standard age distribution P(t) (see Appendix) equal to th proportion of person-years exposure that falls in age group t in the exposed cohor then the age-adjusted hazard rates are $I_1^* = 20/13{,}502 = 1.481 \times 10^{-5}$ from (A1 and $I_0^* =$ "expected deaths"$/13{,}502 = 8.96/13{,}502 = 66.4 \times 10^{-5}$ from (A2) Hereafter, we omit the factor 10^{-5} when referring to hazard rates and note that th units are per 10^5 person-years.

The SMR may be calculated as $I_1^*/I_0^* = 20/8.96 = 2.23$, but this is a crud estimate of RR_E, because smoking is not taken into account. Amandus and Wheele [1987] assumed $RR_C = 14$ for smoking, and they assumed that 84% of mine smoked ($\pi_1 = 0.84$), compared with 67% ($\pi_0 = 0.67$) for the general U.S. popula tion. They employed Axelson's [1978] formula, which is equivalent to our equatio 2, to obtain the corrected estimate

$$\text{RR}_\text{E} = 2.23[\{.84 \times 14 + .16\}/\{.67 \times 14 + .33\}]^{-1}$$
$$= 2.23/1.228 = 1.82.$$

As discussed above and in the Appendix, the validity of applying this correction to the SMR, which can be thought of as the ratio I_1^*/I_0^*, depends on the assumptions that π_1, π_0, RR_E, and RR_C do not vary substantially with age, and that the multiplicative model 1 holds for each age group. Under this model, the corrected estimate $\text{RR}_\text{E} = 1.82$ and the value RR_C may be used to reconstruct the underlying age-adjusted rates. For example,

$$I_0^* = 66.4 = I_{01}^* \pi_0 + I_{00}^* (1 - \pi_0) = I_{00}^* (14 \times .67 + .33)$$

implies

$$I_{00}^* = 66.4/9.71 = 6.84.$$

Other terms in Table I are obtained from equation 1. The asterisks denote age-standardized rates based on the age distribution P(t) defined above.

Suppose instead that the additive model 4 holds. The crude risk difference $I_1^* - I_0^* = 148.1 - 66.4 = 81.7$ is adjusted via equation 5 to obtain

$$\text{RD}_\text{E} = 81.7 - (.84 - .67)\text{RD}_\text{C}$$
$$= 81.7 - (.17)(88.9)$$
$$= 66.6.$$

Here we have taken $\text{RD}_\text{C} = 95.74 - 6.84$ to be consistent with $\text{RR}_\text{C} = 14$ in the unexposed U.S. population (see Table I). The remaining age-adjusted rates are obtained from the additive model 4 (see Table I). The crude estimate of excess relative risk $\text{ERR}_\text{E}^\text{CRUDE} = (I_1^* - I_0^*)/I_0^* = (148.1 - 66.4)/66.4 = 1.23$, which may be corrected according to equation 9 with $\text{ERR}_\text{C} = \text{RR}_\text{C} - 1 = 13$ to obtain

TABLE I. Crude and Corrected Lung Cancer Mortality Rates Per 10⁵ Person-Years†

	Miners (exposed)	U.S. population (unexposed)
Age-adjusted hazard rates	$I_1^* = 148.1$	$I_0^* = 66.4$
Corrected hazards under the multiplicative model		
Smokers	$I_{11}^* = 174.0$	$I_{01}^* = 95.7$
Nonsmokers	$I_{10}^* = 12.4$	$I_{00}^* = 6.84$
Corrected hazards under the additive model		
Smokers	$I_{11}^* = 162.4$	$I_{01}^* = 95.7$
Nonsmokers	$I_{10}^* = 73.5$	$I_{00}^* = 6.84$

†Asterisks indicate that these are direct standardized rates with respect to the age distribution P(t) that equals the proportion of person-years follow-up in age group t in the exposed study cohort.

$$\mathrm{ERR_E} = 1.232(13 \times .67 + 1) - (.84 - .67)(13)$$
$$= 9.75,$$

which agrees with $(I^*_{10} - I^*_{00})/I^*_{00}$ and $(I^*_{11} - I^*_{01})/I^*_{00}$ under the additive model in Table I.

We have used the estimates of I^*_{11}, I^*_{10}, I^*_{01}, and I^*_{00}, calculated under the multiplicative and additive models respectively, to study the effect of the choice of the risk model on indirect corrections for $\mathrm{RR_E^{CRUDE}}$, $\mathrm{RD_E^{CRUDE}}$, and $\mathrm{ERR_E^{CRUDE}}$ (Table II). Note first that the corrected $\mathrm{RR_E}$ is 1.82 for smokers and nonsmokers under the multiplicative model, but that the corrected $\mathrm{RR_E}$ is 1.70 for smokers and 10.7 for nonsmokers under the additive model. Thus the choice of the risk model has a great impact on the final estimates of relative hazard. Likewise, corrected values of $\mathrm{RD_E}$ and $\mathrm{ERR_E}$ are the same for smokers and nonsmokers under the additive model, whereas $\mathrm{RD_E} = 78.5$ and 5.59 respectively for smokers and nonsmokers under the multiplicative model, and corresponding values for $\mathrm{ERR_E}$ are 11.5 and 0.817. Again, the choice of the risk model has a great impact on "corrections" produced. Unfortunately, without measurements on smoking status for each study participant, we cannot check the risk models to see if either is appropriate. At best, we might rely on knowledge of the joint action of smoking and vermiculite obtained from another independent study to justify our choice of a risk model and consequent correction procedures.

MODEL SELECTION AND INDIRECT CORRECTIONS

The choice of an appropriate summary measure of risk from exposure depends on the underlying model of joint action of exposure and confounder (Table II). This is true whether or not data on confounders are available. If the model is additive, the assumption of a common relative hazard, $\mathrm{RR_E}$, may mask important variations in relative risk among subsets of the population. For example, the relative risks of exposure differ dramatically for smokers and nonsmokers under an additive model in Table II. If the model is multiplicative, the calculation of a common risk difference, $\mathrm{RD_E}$, or common excess relative risk, $\mathrm{ERR_E}$, may mask important differences in these quantities, as illustrated in the comparison between smokers and nonsmokers in

TABLE II. Crude and Corrected Measures of Exposure Effect for the Additive and Multiplicative Models†

	$\mathrm{RR_E}$	$\mathrm{RD_E}(\times 10^5)$	$\mathrm{ERR_E}$
Crude	$I^*_1/I^*_0 = 2.23$	$I^*_1 - I^*_0 = 81.7$	$(I^*_{10})/I^*_0 = 1.23$
Corrected with the multiplicative model			
Smokers	1.82	78.5	11.5
Nonsmokers	1.82	5.59	0.817
Corrected with the additive model			
Smokers	1.70	66.6	9.75
Nonsmokers	10.7	66.6	9.75

†Asterisks indicate that these are direct standardized rates with respect to the age distribution P(t) that equals the proportion of person-years follow-up in age group t in the exposed study cohort.

Table II. Indeed, for purposes of assessment of public health risk, the most interesting feature of Table II is the small risk difference for nonsmokers and the large risk difference for smokers under the multiplicative model.

Unfortunately, studies in which indirect corrections are needed do not provide the necessary data on exposures and confounders to test models of joint action. To facilitate the selection of appropriate models, it would be helpful if those studies that do provide information on exposures and confounders would report more complete assessment of joint risk than is customary. If the literature provides little guidance on the choice of a model for joint action, it would be prudent to carry out indirect corrections both for the relative risk under a multiplicative model and for the risk difference or excess relative risk under an additive model. This procedure would allow one to gauge the importance of confounding for either model and to judge whether the same qualitative conclusions would be reached under either model.

Indirect correction produces quite different effects on the excess relative risk than on other measures of risk (Table II). Under the multiplicative model, indirect correction results in a decrease in the relative risk of (2.23 − 1.82)/2.23 = 18% and, under the additive model, indirect correction leads to a decrease of (81.7 − 66.6)/81.7 = 18% for the risk difference, whereas indirect correction leads to an *increase* of (9.75 − 1.23)/1.23 = 693% for the excess relative risk. One can calculate the quantities, (corrected value − crude value)/(crude value) respectively for RR_E, RD_E, and ERR_E as

$$(\pi_0 - \pi_1)(RR_C - 1)\{\pi_1(RR_C - 1) + 1\}^{-1} \quad (10)$$

$$RD_E\{I_1 - I_0\}^{-1} - 1 \quad (11)$$

and

$$RD_E(I_0/I_{00})\{I_1 - I_0\}^{-1} - 1. \quad (12)$$

Recall that $RR_C = I_{01}/I_{00}$, $RD_E = I_{10} - I_{00}$, $I_1 = \pi_1 I_{11} + (1 - \pi_1)I_{10}$, and $I_0 = \pi_0 I_{01} + (1 - \pi_0)I_{00}$. Suppose, as in the section above, that $RR_C > 1$, $RD_E > 1$, and $\pi_1 > \pi_0$. Then equations 10 and 11 must be negative, whereas equation 12 may be either positive or negative, according as $\pi_0 I_{10} - \pi_1 I_{00}$ is positive or negative. Moreover, the quantity 12 may be quite large when I_{00} is much smaller than I_0. For instance, in the example in the section above, the high frequency of smokers in the unexposed population ($\pi_0 = 0.67$) and the high risk from smoking cause $I_0^* = 66.4$, compared with only 6.84 for I_{00}^*. These properties imply that indirect correction may have a much more dramatic impact on the excess relative risk than on the relative risk or risk difference. If cohort data were available, the risk difference would therefore seem to be a better summary statistic than the excess relative risk under additive risk models. If only case-control data are available, one is forced to rely on the excess relative risk under the additive model.

DISCUSSION

We have noted that indirect corrections are linked to models of joint risk for exposures and confounders and stressed the need to explore both the additive and multiplicative models when substantial uncertainty exists. For most purposes this would seem to be sufficient, although other families of models of joint action that include the multiplicative and additive models as special cases as well as intermediate and more extreme models have been studied [Thomas, 1981; Breslow and Storer, 1985; Guerrero and Johnson, 1982; Moolgavkar and Venzon, 1987; Lubin and Gaffey, 1988].

A second warning is needed for studies of composite populations. One must assume a constant effect of exposure and confounder, whether on the multiplicative or additive scale, and constant confounder fractions π_1 and π_0 for all age groups in order to carry out a simple adjustment via equations 2, 5, and 9. This assumption may often be sufficiently accurate for practical calculations. If more specific information is available on the variation with age of the effect of the confounder and of π_1 and π_0, then individual corrections should be performed for each age group and a weighted average of the corrected exposure effects should be computed.

Another practical difficulty in employing indirect adjustments is the need for information on the effect of confounding in the unexposed population and on π_1 and π_0. Fortunately, much is known about the effects of smoking in an unexposed population, and estimates of π_1 and π_0 may be obtained from tables in Brackbill et al. [1988] and Stellman et al. [1988]. In cases where available data on risks in the unexposed population and on π_1 and π_0 are less certain, one should study ranges of plausible values. Extensions to confounders at multiple levels are straightforward, but often knowledge of the confounder effects at various confounder levels and of the distributions of confounder levels for exposed and unexposed populations would not be available.

Despite these problems, we believe that correction for smoking effects should be examined routinely, but with care, in occupational studies that do not obtain smoking information. These correction formulas can enhance the validity and credibility of these studies, particularly when the corrections are small and yield qualitatively similar results for both the additive and multiplicative models.

ACKNOWLEDGMENTS

We wish to thank Drs. O. Axelson and K. Steenland for helpful comments and Mrs. Jennifer Donaldson for typing the manuscript.

REFERENCES

Amandus HE, Wheeler R (1987): The morbidity and mortality of vermiculite miners and millers exposed to tremolite-actinolite. Am J Ind Med 11:15–26.

Axelson O (1978): Aspects of confounding in occupational health epidemiology. Scand J Work Environ Health 4:98–102.

Axelson O, Steenland K (1988): Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Ind Med 13:105–118.

Brackbill R, Frazier T, Shilling S (1988): Smoking characteristics of U.S. workers based on the 1978–1980 National Health Interview Survey. Am J Ind Med 13:5–41.

Breslow NE, Storer BI (1985): General relative risk functions for case-control studies. Am J Epidemiol 122:149–162.

Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959): Smoking and lung cancer: Recent evidence and a discussion of some questions. J Natl Cancer Inst 22:173–203.

Guerrero VM, Johnson RA (1982): Use of the Box-Cox transformation with binary response models. Biometrika 69:309–314.

Lubin JH, Gaffey W (1988): Relative risk models for assessing the joint effects of multiple factors. Am J Ind Med 13:149–167.

Miettinen O (1972): Components of the crude risk ratio. Am J Epidemiol 96:168–172.

Miettinen O (1976): Estimability and estimation in case referent studies. Am J Epidemiol 103:226–235.

Moolgavkar SH, Venzon DJ (1987): General relative risk regression models for matched case-control studies. Am J Epidemiol 126(5):949–961.

Prentice RL, Breslow NE (1978): Retrospective studies and failure time models. Biometrika 65:153–158.

Rothman KJ (1986): Modern Epidemiology. Boston: Little, Brown and Company.

Schlesselman JJ (1978): Assessing effects of confounding variables. Am J Epidemiol 108:3–8.

Stellman SD, Boffetta P, Garfinkel MA (1988): Smoking habits of 800,000 American men and women in relation to their occupations. Am J Ind Med 13:43–58.

Thomas DC (1981): General relative risk models for survival time and matched case-control analysis. Biometrics 37:673–686.

APPENDIX

Corrections for Composite Populations with Adjustment for Age but not for the Unmeasured Confounder

Let I_{ijt} be the hazard rate for age group t at exposure level i (i = 0,1) and confounder level j (j = 0,1). We assume throughout this appendix that $P(j\,|\,i,t) \equiv \pi(j\,|\,i)$ is independent of t. In the previous notation $\pi_0 = \pi(j = 1\,|\,i = 0)$ and $\pi_1 = \pi(j = 1\,|\,i = 1)$.

First consider the multiplicative model 1. We make the added assumptions that RR_E (equation 3) and RR_C are independent of t. Suppose we perform direct age adjustment with respect to a reference age distribution P(t) to obtain the age adjusted hazard for exposed,

$$I_1^* \equiv \sum_t P(t)\,\{\sum_j I_{1jt}\,\pi(j\,|\,1)\}. \tag{A1}$$

The term in curly brackets is the crude rate in age group t, because it is a mixture of confounder-specific rates. Thus I_1^* is a crude age-adjusted rate. We similarly define

$$I_0^* = \sum_t P(t)\,\{\sum_j I_{0jt}\,\pi(j\,|\,0)\}. \tag{A2}$$

From the previous assumptions, $BIAS_M$ is independent of t, so that, from equation 2,

$$\begin{aligned} I_1^*/I_0^* &= \sum_t P(t)\,\{(RR_E)\,(BIAS_M)\sum_j I_{0jt}\,\pi(j\,|\,0)\}/I_0^* \\ &= (RR_E)\,(BIAS_M). \end{aligned} \tag{A3}$$

Thus the same factor $BIAS_M$ used for a single age stratum may be used to correct the ratio of two direct age-adjusted hazard rates.

Indirect standardization is a special case of direct standardization when the standard age distribution is taken as that of the study population. With P(t) now chosen to be the age distribution in the exposed study cohort (i = 1), the indirect adjustment for age is precisely

$$SMR = \text{“observed”}/\text{“expected”} = I_1^*/I_0^* = (RR_E)\,(BIAS_M). \tag{A4}$$

Thus the same bias factor may be used. Suppose instead that two study populations, one exposed and one unexposed, are each standardized with respect to age-specific rates from a third general population. Then, the ratio of age-adjusted SMRs reduces to

$$SMR_1/SMR_0 = (RR_E)\,(BIAS_M) \tag{A5}$$

where $RR_E = I_{1jt}/I_{0jt}$ and $BIAS_M$ is given by equation 2 as before. All terms involving the common general population cancel out; in particular, the denominator of $BIAS_M$ in equation 2, which refers to the probability that $j = 1$ in the general population, cancels out.

We now consider the risk difference, RD_E. Let I_1^* and I_0^* be direct age-adjusted rates given by equations A1 and A2, where P(t) is an arbitrary reference age distribution. It follows from the assumptions that π_1, π_0, RD_E, and RD_C are independent of time that

$$\begin{aligned} I_1^* - I_0^* &= \sum_t P(t) \sum_j \{I_{1jt}\pi(j\,|\,1) - I_{0jt}\pi(j\,|\,0)\} \\ &= \sum_t P(t)\{RD_E + BIAS_A\} \\ &= RD_E + BIAS_A. \end{aligned} \tag{A6}$$

Thus, the same factor $BIAS_A$ computed in equation 5 may be used to correct the age-adjusted risk difference. In contrast to equation A5, indirect standardization for age is not appropriate for computing an age-adjusted risk difference between two study populations, one exposed and one unexposed and each standardized with respect to rates from a third general population, because two different age distributions, one for the exposed study cohort and one for the unexposed study cohort, come into play.

Similar results apply for the excess relative risk. Under the assumptions that π_1, π_0, ERR_E, and ERR_C are independent of age, the crude excess relative risk is

$$\begin{aligned} (I_1^* - I_0^*)/I_0^* &= \frac{\sum_t P(t)\{ERR_E + ERR_C(\pi_1 - \pi_0)\}}{\sum_t P(t)(ERR_C\pi_0 + 1)} \\ &= \{ERR_E + ERR_C(\pi_1 - \pi_0)\}/(ERR_C\pi_0 + 1), \end{aligned} \tag{A7}$$

in agreement with equation 8. As before, $I_1^*$and I_0^* are direct age-adjusted rates. It follows from equations 8 and A7 that the same corrections may be applied to the crude excess relative risk obtained from age-adjusted rates as to crude age-specific excess relative risks. Note, however, that one must use direct age adjustment for these results to hold. The factor $\sum_t P(t)$ does not cancel out if indirect adjustment is used, because the different age distributions for the exposed and unexposed cohorts come into play.